BUILDING SERVICES ENGINEERING RESEARCH & TECHNOLOGY

Building Serv. Eng. Res. Technol.
0(0) 1–25

DOI: 10.1177/0143624416684641
journals.sagepub.com/home/bse
SAGE

The building performance gap: Are modellers literate?

Salah Imam[1], David A Coley[1] and Ian Walker[2]

Abstract

One of the most discussed issues in the design community is the performance gap. In this research, we investigate for the first time whether part of the gap might be caused by the modelling literacy of design teams. A total of 108 building modellers were asked to comment on the importance of obtaining and using accurate values for 21 common modelling input variables, from U-values to occupancy schedules when using dynamic simulation to estimate annual energy demand. The questioning was based on a real building for which high-resolution energy, occupancy and temperature data were recorded. A sensitivity analysis was then conducted using a model of the building (based on the measured data) by perturbing one parameter in each simulation. The effect of each perturbation on the annual energy consumption given by the model was found and a ranked list generated. The order of this list was then compared to that given by the modellers for the same changes in the parameters. A correlation analysis indicated little correlation between which variables were thought to be important by the modellers and which proved to be objectively important. k-means cluster analysis identified subgroups of modellers and showed that 25% of the people tested were making judgements that appeared worse than a person responding at random. Follow-up checks showed that higher level qualifications, or having many years of experience in modelling, did not improve the accuracy of people's predictions. In addition, there was no correlation between modellers, with many ranking some parameters as important that others thought irrelevant. Using a three-part definition of literacy, it is concluded that this sample of modellers, and by implication the population of building modellers, cannot be considered modelling literate. This indicates a new cause of the performance gap. The results suggest a need and an opportunity for both industry and universities to increase their efforts with respect to building physics education, and if this is done, a part of the performance gap could be rapidly closed.

Practical application: In any commercial simulation, the modeller will have to decide which parameters must be included and which might be ignored due to lack of time and/or data, and how much any approximations might perturb the results. In this paper, the judgment of 108 modellers was compared against each other. The results show that the internal mental models of thermal modellers disagree with one another, and disagree with the results of a validated thermal model. The lessons learnt will be of great utility to modellers, and those educating the next generation of modellers.

[1]Department of Architecture and Civil Engineering, University of Bath, Bath, UK
[2]Department of Psychology, University of Bath, Bath, UK

Corresponding author:
Salah Imam, University of Bath, Claverton Down Rd, Bath, BA2 7AY, UK.
Email: imam.salah@bath.edu

Keywords
Literacy, building modellers, simulation, performance gap, input variables

Introduction

Many policies and actions are being implemented by governments with the aim of reducing greenhouse gas emissions. In developed countries, buildings commonly account for up to 40% of such emissions,[1] making them a clear focus. Unfortunately, there is a proven gap between the energy use predicted by models of buildings used to aid their design, or ensure compliance with national building codes, and the monitored energy consumption of the buildings once built. Many researchers claim that the measured energy consumption is frequently twice or more than that of the design stage prediction,[2–4] and although many studies have explored the performance gap from various perspectives, such as the role of poor workmanship or occupants' behaviour, the literacy of building energy modellers is rarely questioned. In addition, the literature indicates that in general, professionals (architects, engineers, sustainability experts, etc.) do not tend to criticize themselves and thus a culturally embedded lack of reflection might contribute to the performance gap.[2–5]

Modelling professionals are limited in the time they can apportion to any project and hence need accurate inbuilt knowledge of the impact that modelling any element of the building in less than ideal detail might have; for example, the impact of missing out a thermal bridge. The basis for these judgment calls might be in part based on experience, but it is likely to also be embedded within an organisation, or just commonly accepted within the modelling community.[6,7] Professionals in general are known to be open to change if evidence is presented,[8] and this paper attempts to provide this evidence in a robust way, by asking the question, how accurate in general are such professionals' judgments?

Background

Literacy

The United Nations Educational, Scientific and Cultural Organization (UNESCO) defines literacy as the 'ability to identify, understand, interpret, create, communicate and compute, using printed and written materials associated with varying contexts. Literacy involves a continuum of learning in enabling individuals to achieve their goals, and to develop their knowledge and potential'.[9] Some have argued that this definition of literacy should be expanded to include the capability to use computerized tools efficiently and correctly.[10]

There is no single method to monitor and measure literacy levels, but there are various methodologies that can be followed depending on the aim of the study. According to UNESCO,

> typically countries measure literacy levels by undertaking self-assessment questionnaires and/or by means of a proxy variable utilizing the number of years of primary schooling (i.e., 6 or 8 years of primary schooling equals a literate person), typically literacy rates are assigned so that people over 15 years of age are designated as literate.[11]

Unfortunately, this does not give a robust method for measuring literacy levels in other settings. An alternative is to use tailored questioning to assess literacy.

There are many ways one might define literacy with respect to building physics and thermal modelling, and we are after a measure which is more independent and about modelling in general, not about a certain simulation package or method. The assessment method also needs to provide a numeric result or a ranking in order that a quantitative assessment of literacy can be made. Here, we suggest a suitable requirement for literacy within a population is that we might expect that when given a real project the population of modellers should: (1) approximately agree on the important parameters that need to be included in the model; (2) approximately agree on the rank order of the importance of a list of possible input parameters; (3) that their rank ordering of the impact of given changes (perturbations) to the values of these parameters should approximately agree with that given by a sensitivity analysis of the parameters within a common thermal model.

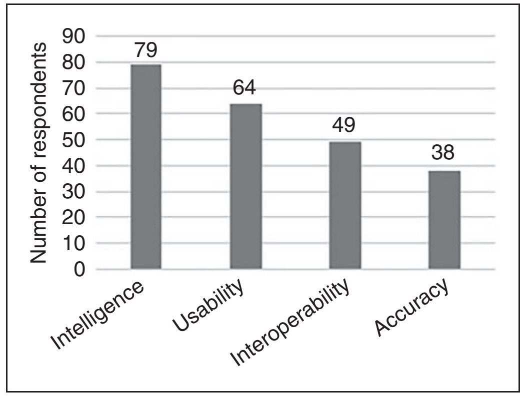

Figure 1. Architects' ranking of the importance of simulation tool features (data from Attia et al.[12]).

Building energy modelling

Researchers have noted the influence that the building design industry has had on building performance simulation (BPS) tools and vice versa. This development has meant more complexity without evidence that the complexity is manageable by all professionals.[12] For example, architects are regularly using BPS tools, despite them being described as generalists.[13–18]

Many studies have highlighted that most tools available are inadequate to deal with early design stages. Furthermore, they are not user friendly.[19–22] The building simulation industry became aware of this and tried to tackle it by producing more friendly interfaces. However, many barriers still exist in using these tools.[12]

It has been argued that the most important capabilities of these tools are *usability*, *computing ability*, *data-exchange* and *database support*.[23] Researchers have also stated the importance of what they called 'functional criteria' of BPS tools, which again addresses the question of usability.[15] Despite researchers' concerns about usability, tools over the years have become more and more complex.

Attia et al.[12] performed a survey with approximately 150 architects, with the aim of ranking the selection criteria of BPS tools according to their importance from the user point of view. The results showed that model intelligence had the highest priority (Figure 1). (The study defined model intelligence 'as the ability to advise the user with design optimisation options based on a range of early stage input'.) Accuracy was considered the least important.[12]

The performance gap

The literature indicates that a disconnect between modelled and actual performance can occur in each of the three broad stages of: design, construction and operation.[3,24]

The design gap.

Many studies have concluded that the design phase is a frequent cause of the gap.[4,24] Reasons include misunderstanding of the design performance targets between design team and client, or even between the design team members.[25] In addition, De Wilde[4] pointed out that even if the design itself is properly outlined, underperformance can still occur if the design team did not take into consideration buildability, simplicity or the construction sequence. Other papers have focused on issues with the specification of advanced systems and technologies due to the level of complexity of the system and its controls.

The Zero Carbon Hub[5] report 'Closing the Gap' observed that professionals have a limited understanding of the impact of their design decisions on actual energy performance. For example, how much might improving the U-value by 10% reduce heating energy consumption in a particular climate? But this observation was not based on a quantitative assessment, and is hence possibly questionable. Knowledge of the impact of uncertainties in the design stage is another level of literacy that is understudied, and it is unknown if practitioners gain the required knowledge to address this after many years of experience or not, but given that few buildings are monitored after construction by their designers, this seems unlikely.

It is known that incorrect use of simulation tools will result in unreliable predictions at the design stage, which will lead to the gap later on, and therefore, the user has to have a minimum level of knowledge and skills to be able to use these tools properly.[26] De Wilde[4] pointed out that the required knowledge includes the ability to define correct input data within the model. Nevertheless, even with an experienced user, many predictions will still be inconsistent and lacking in certain areas, mainly arising from issues of uncertainties such as occupancy behaviour and weather data.[2]

The construction gap. Another issue that can cause a performance gap is the construction process. Many studies, including industry reports and papers analysing various scales and types of case studies, have pointed out that the onsite construction quality often does not agree with design specifications. More particularly, there is a lack of attention to aspects related to insulation and airtightness.[2,4,27] In many cases, both builders and engineers are responsible for the resultant discrepancy in buildings performance, but studies have not been able to identify nor quantify the exact source of the gap.

The operational gap. A building's operational stage is repeatedly cited to be a major reason for discrepancy with the design stage predictions. More particularly, studies often put the blame on occupants' behaviour.[2,4,28,29] It is suggested that by using proper post occupancy evaluation data, more knowledgeable design stage assumptions might be possible in future and hence reduce this contribution to the gap.[2] However, such data are rarely collected.

Building simulation modelling

Case study building

The particular building chosen in this study was a typical UK semi-detached house, which was recently renovated to meet the L1B requirements (essentially an upgrade to the relevant building codes). Such a building, rather than for example a large office block, was chosen deliberately to reduce the complexity of the situation and hence improve the accuracy of the human judgements. The building was modelled in detail using IES and the model was validated using measured hourly gas consumption, electricity use, occupancy and indoor temperatures.

Modelling approach and limitations

Weather input data. Observed weather was recorded for the project from a weather station approximately 3 miles from the house. This gave, dry bulb temperature, wet bulb temperature, relative humidity, wind direction and wind speed. Radiation data were taken from the World Meteorological Organization's website for Camborne (the closest available location) with similar climate characteristics and hourly measured weather data (2004–2014). Other data were from the EPW for London. As the paper only examines changes to the annual energy consumption, given by perturbations in the modelling variables, not the consumption itself, minor inaccuracies in the weather files are likely to have little effect on the results.

Heating use. System use was determined based on observations of measured energy consumption, and indoor temperature variations for each space. The heating set-point (21°C) was based on the measured indoor temperature.

Building geometry. Internal and external dimensions and openings of the case study building were modelled carefully using to-scale drawings.

Surroundings. The surrounding environment of neighbouring buildings was modelled in detail, as this provides extensive shading. The case study building has no external self-shading except for 200 mm extrusions above doors, a 100 mm extended roof perimeter and a 100 mm recession around windows and doors – all were included in the model.

Glazing ratio. The plans gave a glazing ratio of 25% and 21.8% on south and north facades, respectively. The east façade contains only one window, representing 2.3% of the area. Doors were 1.6 m^2 in area (solid doors with no glazing).

Natural ventilation and occupancy. Modelling natural ventilation depends on assumptions, for example, it is highly unlikely a modeller can accurately determine when and which windows will be opened, and for what length of time. Therefore, modellers usually use assumptions that are under-descriptive of the actual behaviour of occupants. For the purposes of this research, and starting from reasonable assumptions, the ventilation was adjusted to give a high correlation between measured and simulated heating energy demand and temperature (measured on an hourly basis). This means the model is much more accurate than that normally created by a design team.

Building's envelope. The air permeability of the building envelope was set as 10 $m^3/h/m^2$ at 50 Pa in order to comply with the standard set by the building code (Part L). Any error here being accounted for in the way natural ventilation was modelled (see paragraph above). U-values were as detailed in Table 1.

Table 1. U-values of case study building.

Element	Modelled U-values (W/m^2K)
External walls	0.35
Pitched Roof	0.26
Floors	0.25
Windows	1.6
Doors	1.8
Internal walls	1.8
Internal floor/ceiling	1.0

Internal heat gains. The sensible gains from people were set to 75 W/person in accordance with the ASHRAE handbook (2013).[30] A maximum of four people were assumed to be in the house, with occupancy linked to the measured occupancy profiles of each space. Gains from lighting were controlled based on the illuminance level required for each space and occupancy period. Finally, internal gains from equipment and cooking were assumed as an average based on the ASHRAE handbook (2013). The appliances were linked to occupancy profiles of each space in order to provide the measured average value of consumption. This action was performed with an understanding that not all appliances are linked to occupancy profiles, for example fridges.

Model validation: Simulation vs. measured data

In order to validate the model, one year of detailed gas consumption and indoor temperature monitoring was obtained and correlated with the simulated case study results. The data were compared on hourly intervals across the entire year. The correlation between measured monthly gas consumption and the simulated model gives an R^2 of 0.93 (Figure 2), with the hourly correlation also being good (Figures 3 and 4). As illustrated in Figures 5 and 6, a strong correlation is found between both peak and average indoor temperatures in all spaces.

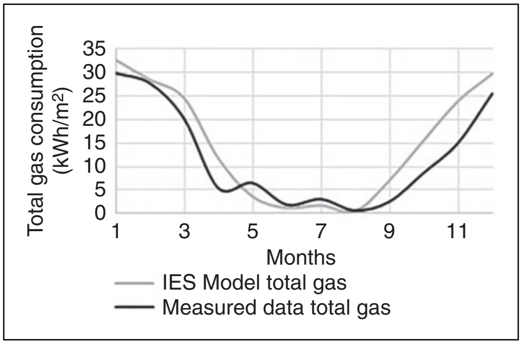

Figure 2. Monthly correlation between measured and simulated gas consumption ($R^2 = 0.93$).

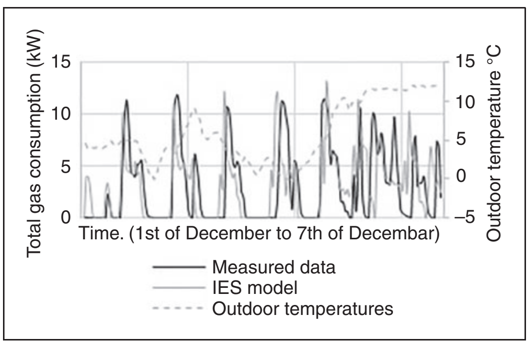

Figure 3. Simulated and measured hourly gas consumption for a week in December ($R^2 = 0.73$).

The model can thus be considered as validated. Table 2 and Figures 7 and 8 show the perturbations introduced and their impact on the model.

Survey

Method

Survey design. From a psychological perspective,

> A person's perception of how a system operates is often referred to as a mental model. This might come from educated understandings via literature and mentorships or simply from practical experimentation with the controls – and in both cases their mental model might or might not be accurate.[31]

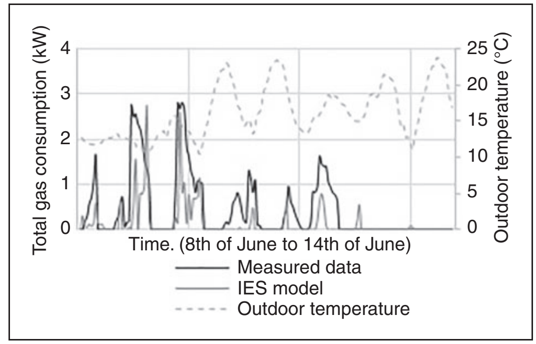

Figure 4. Simulated and measured hourly gas consumption for a week in June ($R^2 = 0.59$).

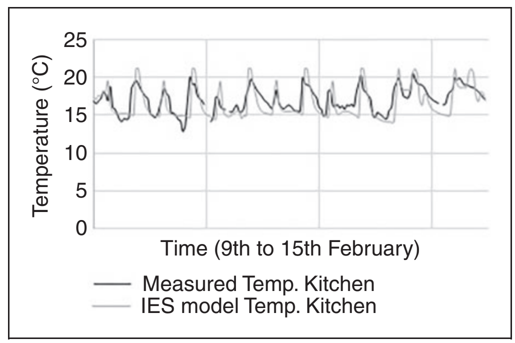

Figure 5. Plot of both simulated and measured indoor temperatures for the kitchen space for a week in February ($R^2 = 0.61$).

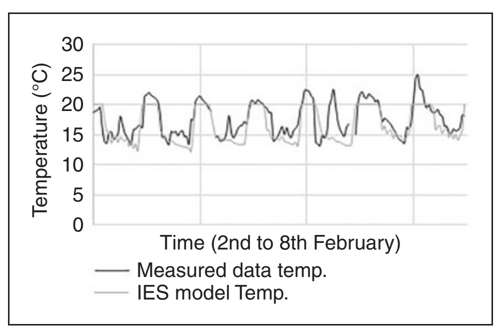

Figure 6. Plot of simulated and measured indoor temperatures for a bedroom for a week in February ($R^2 = 0.63$).

Within this context, the survey conducted in this research aims to reveal the energy modelling 'mental models' of professionals in the construction industry. This was done by asking questions

Table 2. Perturbations performed on each input parameter.

Input parameter	Base value	Altered value	Scale of alteration
Glazing ratio	17.3%	19 %	10% Greater than actual and modelled ratio
Installed window U-value	1.6 W/m^{2}K	1.92 W/m^{2}K	20% Greater than installed and modelled value
Walls U-value	0.35 W/m^{2}K	0.42 W/m^{2}K	20% Greater than installed and modelled value
Occupancy period	13 hr/day	16.25 h/day	25% Greater than the average measured and modelled period per day
Airtightness	0.25 ach	0.3 ach	20% Greater than the assumed and modelled value
Roof U-value	0.26 W/m^{2}K	0.31 W/m^{2}K	20% Greater than installed and modelled value
Thermal bridging	10% Increase in each element U-value	Thermal bridges ignored	Ignoring thermal bridging
Winter indoor temperature set-point	21°C	19°C	The modelled value being 2°C lower than reality
Natural ventilation	MacroFlo profiles	Constant airflow at 1 ach	Assuming the air flow is constant at 1 ach when occupied, against the base case of assuming windows are open during occupied period, if ($T_{in} > 25$°C, RH >65% or CO_2 concentration 1000 ppm)
Ground floor U-value	0.25 W/m^{2}K	0.3 W/m^{2}K	20% Greater than installed and modelled value
Building geometry	39.5 m^2	32 m^2	Using internal dimensions for the building rather than external
Ventilation rate	1 ach	1.1 ach	10% increase
Shading from surroundings	Modelled surroundings	Ignore their effect	Ignoring shading from the surrounding homes etc.
Windows recession	100 mm	200 mm	Assuming windows recessed 100 mm further into the building
The position of windows in walls	Base model position	0.5 m downwards	Assuming a 0.5 m vertical shift down from the actual position on each facade
Density of block used as inner leaf of wall	1.40 Tonne/m^3	1.54 Tonne/m^3	20% greater than installed and modelled value
Internal gains from appliances and lighting	52.8 W/m^2	58.0 W/m^2	10% greater than installed and modelled value
External doors opening	10 Openings/day	Continuously closed	Ignoring the fact that the external doors might be opened 10 times a day, each time for 30 s

(continued)

Table 2. Continued.

Input parameter	Base value	Altered value	Scale of alteration
Internal gains from cooking	12 W/m^2	0 W/m^2	Ignoring heat gains from cooking
Thermostat location	Thermostat only in the living room	Thermostat in each space	Assuming thermostats in each room rather than just in one room (modellers often assume the former)
The use of curtains	Used at night	Ignore their effect	Ignoring the use of curtains at night

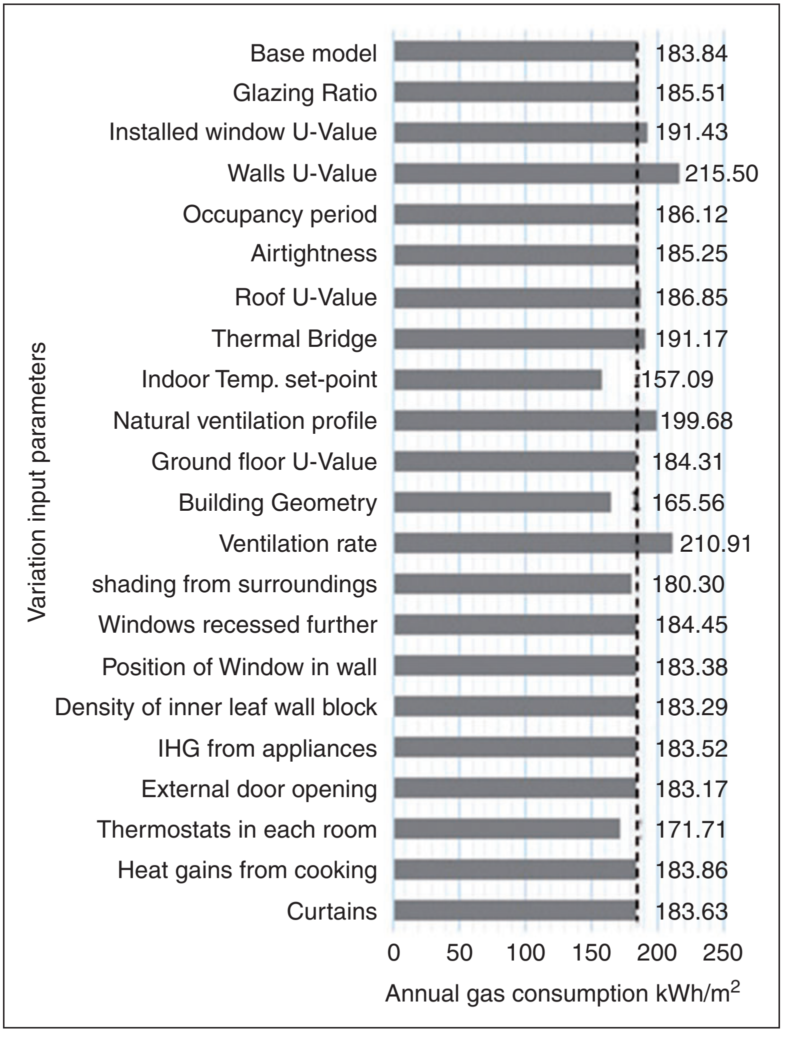

Figure 7. The impact of each perturbation on the annual gas consumption compared with the base model (dashed line).

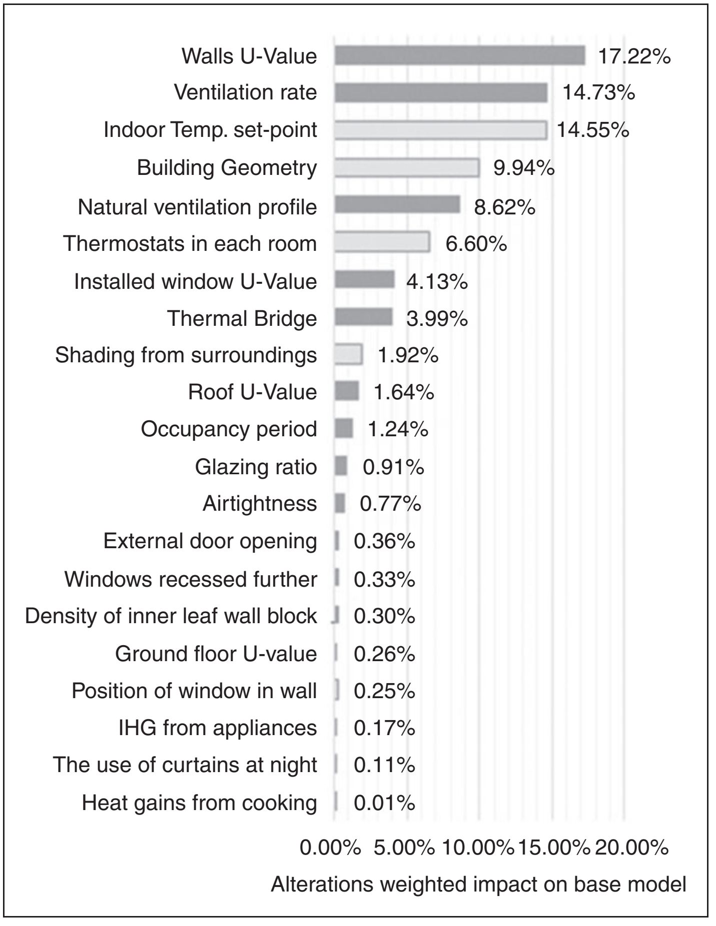

Figure 8. The impact of each perturbation rank ordered in terms of percentage change, with dark bars indicating an increase in consumption and light bars a decrease.

using two standard social science approaches: *the free-form method* and *the given list method*,[30] see Table 3. A detailed description of the building and the surroundings including photographs (see Appendix 1: Online questionnaire) was given to the participants.

Sampling method. The target respondents were chosen from professionals in the construction industry: architects, engineers and energy analysts. All of who made regular use of dynamic thermal models. Random sampling[32,33] was used to generate the population sample.

Participants. Participating employees were from engineering and architectural firms involved in the design process of a range of national and international projects, and included some of the world's largest engineering and architecture practices.

Emails were sent to directors to ask whether it would be possible to visit their firm to ask employees to complete the survey. Many replies welcomed the idea, resulting in 31 respondents. The online questionnaire was also sent directly to professionals drawn from LinkedIn, and respondents were also garnered by posting on online building energy

Table 3. Survey questions and their purpose.

	Survey question(s)	Purposes/aims
Free-form method		
Question 1	List the three most important parameters that if not included or included less accurately in a thermal model of the case study building, might affect the annual heating demand significantly.	To discuss any common input parameters that participants might consider have a significant impact on the annual heating demand.
Question 2	List three parameters that you might not normally include, as they do not have a great impact on the annual heating demand.	To encourage participants to include input parameters that they might not normally consider. Hence, parameters not included in their answers will more likely not used by participants in actual projects.
Question 3	List any other parameters that you might include in a thermal model of the case study building and might have a moderate effect on the annual heating demand.	To give participants the chance to add any other input parameters that they might sometimes include in a thermal model of the case study building.
Structure concept	■ Not providing users with a list of parameters – at this stage – was intentional, so as to not attract them to certain input parameters that need to be included in the model. ■ Clarify what participants do take or do not take into consideration in a thermal model of the case study building and to identify their natural thoughts regarding the modelling stage assumptions. ■ Dividing this section into three questions was to limit the answers to three to five options, making it easier for participants to understand and respond correctly (Holt and Walker[31]).	
Given list of input-parameters method		
Question 1	Rate the list of parameters provided in the survey based on your judgement of impact on annual heating demand due to variations applied to each parameter (Table 2).	■ Identify the perception of the design team of potential errors due to some parameters and their effect on the annual heating energy demand. ■ The answers to this question were obtained in the form of a 'ranked list' and compared with the 'accurate ranking 'obtained from the validated simulation model. ■ This comparison set forms the base for evaluating their modelling literacy.
Notes	■ The details of the case study building were given to participants, as shown in Appendix 1. ■ Once participants proceeded from the 'free-form' question to the 'given list' question, they were not able to return back and edit their responses. Hence, the case study description was repeated to be accessible while answering both questions. ■ The 'error factors' applied to each input-parameter were assumed to be due to lack of knowledge in the design stage or poor workmanship on-site.	

modelling groups, resulting in an additional 77 respondents.

The whole process resulted in 108 participants who completed the survey; a further 12 participants failed to fully complete it. Questionnaire results were anonymous, and the names of the firms participating in the survey cannot be reported due to confidentiality. Figure 9 shows the nature of

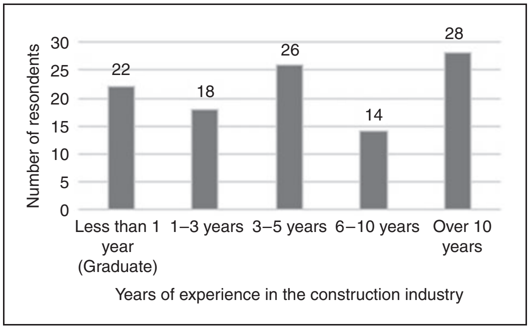

Figure 9. Participants' years of experience in the construction industry.

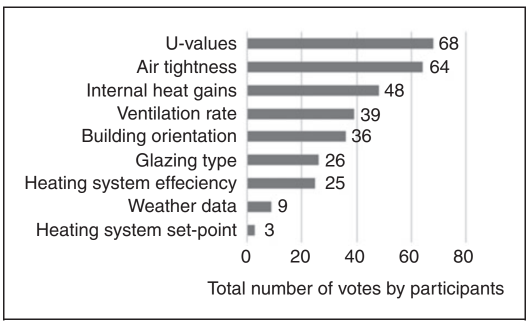

Figure 10. Question 1: Input parameters assumed by participants to have a significant impact on the annual heating demand of the case study building.

the participants, in terms of years of experience within the construction industry. The highest academic degree achieved related to this field was reported as: bachelors (34 participants), masters (66), PhD (8). Eighty per cent of respondents selected IES VE as the simulation software they used for energy analysis.

Results

Free-form method. In this form of the survey, participants were not given a list of parameters to choose from, but asked to separately list parameters they considered highly important, moderately important, or unlikely to be important. Parameters listed by participants are shown in Figures 10 to 12.

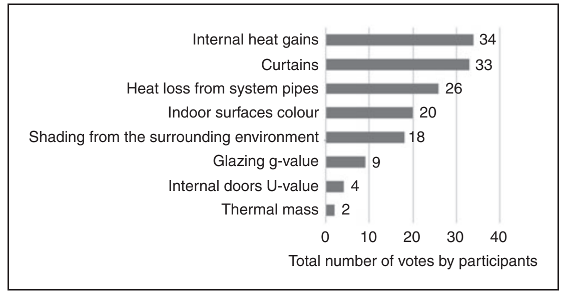

Figure 11. Question 2: Input parameters that participants conclude that they might not normally include in a thermal model of the case study building.

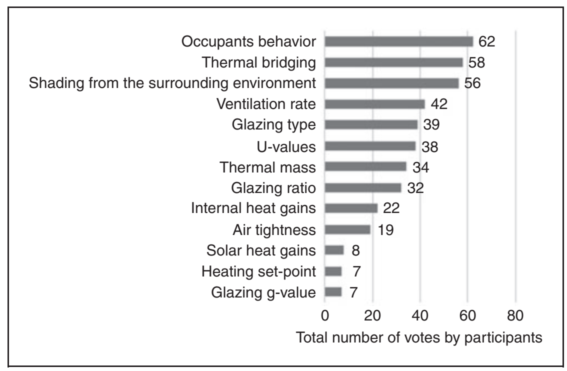

Figure 12. Question 3: Input parameters assumed by participants to have a moderate impact on the annual heating demand of the case study building.

Given list method. For this part of the survey, participants were given a list of 21 input parameters and the perturbations used in the sensitivity analysis (see Tables 2 and 3). Participants were asked to indicate the relative size of impact for each parameter variation on the annual heating demand by scaling them from 1 to 5. The ranking given by the participants is shown in Figure 13. The weighted average for any parameter was calculated as

$$\frac{x_1 w_1 + x_2 w_2 + x_3 w_3 + x_4 w_4 + x_5 w_5}{Total\ number\ of\ respondents} \tag{1}$$

where x is the response (1–5) and w is the response count.

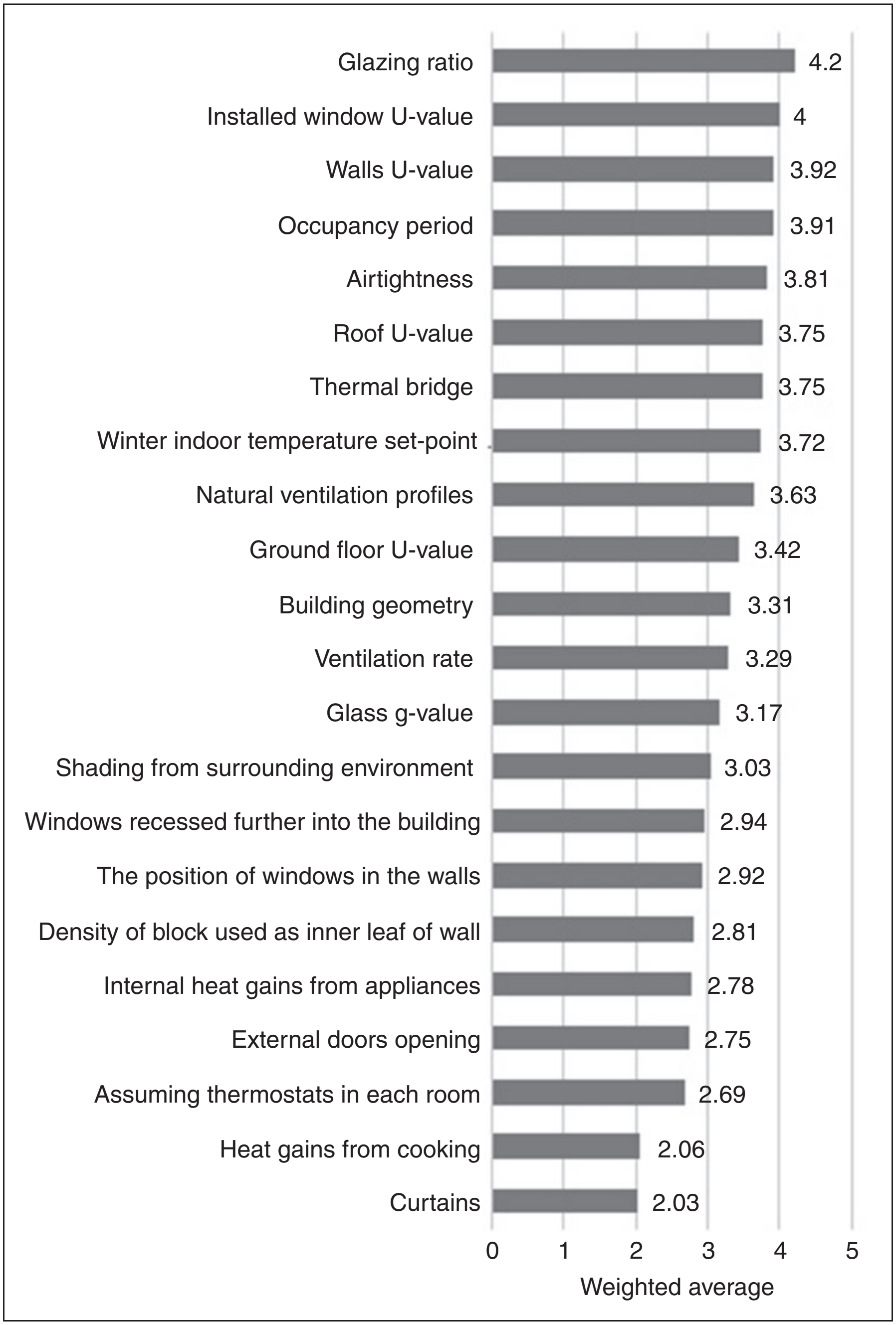

Figure 13. Ranking of the parameters given by participants when asked to indicate on a scale of 1–5 the relative size of the impact for each parameter on the annual heating demand.

Discussion

Un-mentioned parameters.

Re-plotting the free-form results so as to concentrate on parameters not mentioned by one or more individuals provides some surprising results (Figure 14). All parameters were subject to being overlooked except U-values. For example, although 'internal heat gains' was mentioned 104 times out of 108 responses, 34 participants considered it to be the type of parameter that they would not normally include in such a dynamic model. Similarly, 18 participants considered the inclusion of shading from the surrounding environment to not be worth including, whereas 56 respondents highlighted this parameter to be of considerable importance. This is still surprisingly low given that participants were provided with a photo of the surrounding area (see Appendix 1) that shows the building is surrounded by buildings of a similar height.

Comparing and contrasting the results from both survey methods.

Comparing the results obtained from both methods highlights that a parameter's ranking can differ significantly. For example, in the free-form question, 70% of participants did not mention *glazing ratio*, while 42% and 23% did not include *occupancy period* and *airtightness* respectively, whereas the top 5 ranked parameters in the given list question included all three parameters as shown in (Table 4). The full list of responses is given in Appendix 2.

One of the clearest differences between the participants and the ground truth provided by the model is in the impact of changing the glazing ratio (a 10% increase in glazing ratio was presented to the participants and modelled). Although assumed by the participants to be the parameter with the greatest impact, the modelling showed it to only be the 12th and giving an increase of only 0.91% in heating energy use (183.84 to 185.51 kWh/m^2/year). Similarly, installed window U-Value was given by the participants as the second most important, whereas, it was the seventh in the simulation model.

For a few cases, the participants and the model are in better agreement. For example, the impact of changing the wall U-value was voted by the survey as third, which is relatively close to the finding of the simulation study, which placed it first, with an increase of 17.22% in heating energy use. This outcome is probably logical, because of the large surface area of this element and the relatively large perturbation assumed (20%). Ignoring the use of curtains at night, ignoring the internal heat gains due to cooking and a 10% increase in heat gains due to appliances also showed agreement between the participants and the model. All are viewed by the participants and validated by simulation as being of little impact, securing the last five slots in the ranking of both the survey and the simulation model. However, in the case of indoor temperature set-point being reduced by 2°C, the survey gave a rank of eighth, yet the simulation model shows it to be the third; with gas consumption decreasing from the base case by 14.55%.

As discussed earlier, the results from the survey participants are on a scale of 1–5 scale; however, the ranking produced by the simulation model is on a scale of 1–21, making a numeric comparison between the survey results and the model difficult. To analyse the findings further, the survey responses were ranked using equation (1), i.e. ranked according to their mean score, placing them in a ranked list of 21 members. It is clear that there is a large variability in the survey responses and, the mean ranking given by the survey is far from that given by the model, with a Spearman ranking of 0.43 and an R^2 value of 0.28 (Figure 15). This suggests no correlation between the thoughts of designers and the modelled results, and indicates that, when measured in this way, modelling literacy (as defined earlier) may not be high in the participants.

Cluster analysis.

Having shown no overall correlation between the results from the participants and the predictions of the model, it is worth asking if any subpopulations perform

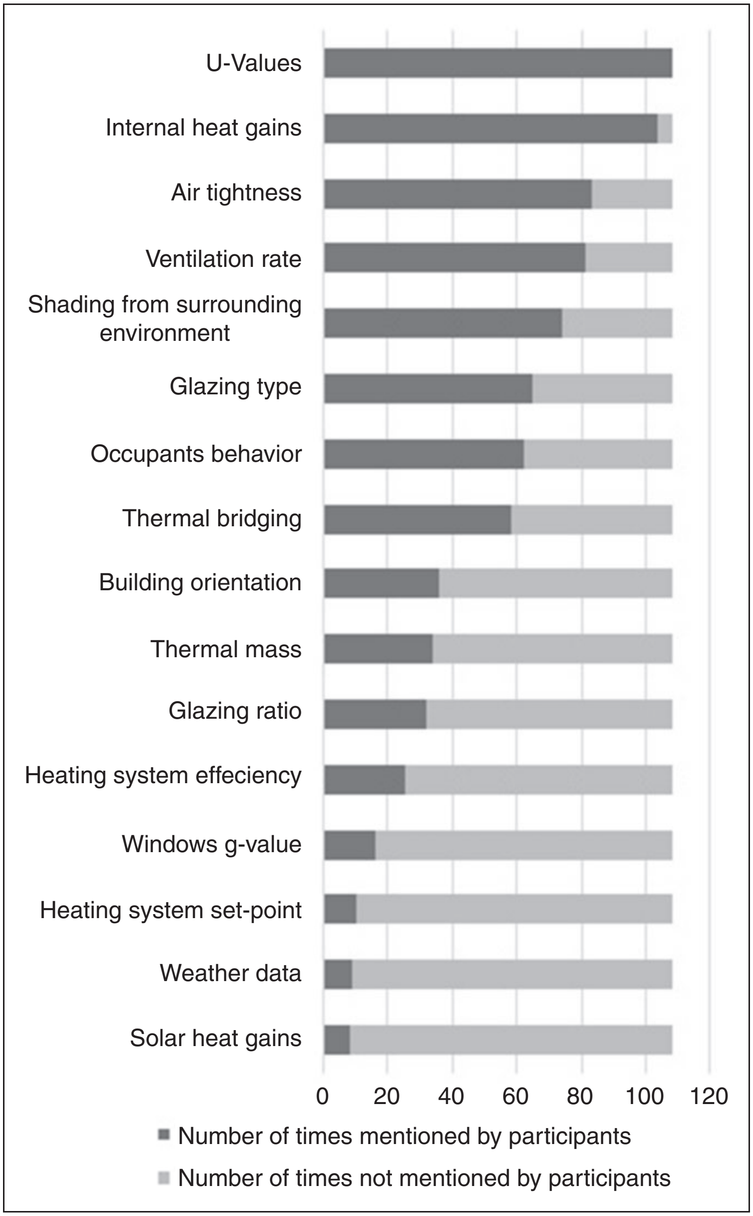

Figure 14. The most impactful input parameters mentioned by participants in the free-form question, highlighting the number of times each parameter, was not mentioned.

better than the average. Normal correlation is not strictly valid for ordered rating categories, so this is best done by looking at each participant's weighted kappa value, κ. This is a measure of agreement between any two sets of numbers that form discrete ordered categories. A person scoring $\kappa = +1.00$ would be rating each item with exactly the same category score

Table 4. Comparison between the Top 5 ranked input parameters in the 'given list' question and the number of times participants did not mention these parameters in the 'free-form' question.

Given list method	Number of participants who did not mention this parameter (total of 108 participants)
Top 5 ranked input parameters	
Glazing ratio	76
Installed window U-value	0
Walls U-value	0
Occupancy period	46
Airtightness	25

as the model did; a person scoring $\kappa = 0.00$ would essentially be responding at random; and a person scoring $\kappa = -1.00$ would be systematically disagreeing with the results of the model (for example, saying the most unimportant parameter perturbations were the most important, and vice versa). To be able to compare the ranks from the survey, which are on a scale of 1–5, with those from the model, which are on a scale of 1–21, the model parameters need to be re-scaled to take values of 1 to 5, so κ can be calculated. The most important perturbation (wall U-value) changes the annual heating energy use by $31.66\,kWh/m^2$; the least important (gains from cooking) by $0.02\,kWh/m^2$.

An initial cluster analysis was used to group the perturbation factors into five groups which could be rated 1–5. *k*-means cluster analysis takes a set of measurements and splits these into *k* groups (with *k* specified by the researcher), whereby the items within each group are as similar to one another as possible and the differences between groups are as large as possible. With $k = 5$, we obtained five groups of factors, ranging from the most important (Walls' U-value, ventilation rate, etc., rated 5) to the least (gains from cooking, curtains, etc., rated 1).

Now that the factors are rated 1–5 both objectively (by this analysis) and subjectively (by the participants), we can calculate the weighted kappa score of each individual and thus have a measure of how skilled they are at rating the perturbations. An agglomerative cluster analysis will then automatically group people based on how similar their kappa values are. This is done by carrying out $N-1$ analysis steps, on each step grouping together the two people (and then groups) with the most similar kappa values. This iterative hierarchical clustering process begins without preconceptions about how many groups of people will be found and identifies any distinct clusters of people with distinct levels of perturbation rating ability; once clusters are thus identified based purely on rating skill, the makeup of each, in terms of education, years of experience or other factors, can be reflected upon. Five clusters are found in this case (Figure 16). Note the emergence of five clusters in this analysis is just coincidence and does not arise from the use of a five-point rating scale.

Figure 16 can be read as follows: starting from the x-axis of 108 participants, the 'stables' link the pairs of individuals with the closest values of κ, then pairs of pairs of similar κ, etc. The dashed red boxes identify the five groups which in κ-space have reasonably similar values; although arguably the two left hand groups (containing the best performing participants) could be combined, as could the two right hand groups (containing the worst performing participants). The people within each of the five groups are similarly skilled to one another at rating the perturbations, and quite different from the people in the other groups. Figure 17 shows that there are three subgroups of people who are better than guessing at the task ($\kappa > 0$) and two groups who are worse than random. The makeup of these groups is discussed in Table 5.

Table 6 further disaggregates the results and shows that the participants with a PhD also had >10 years of experience, so it is unknown whether their poor performance was in anyway connected with their education, rather than

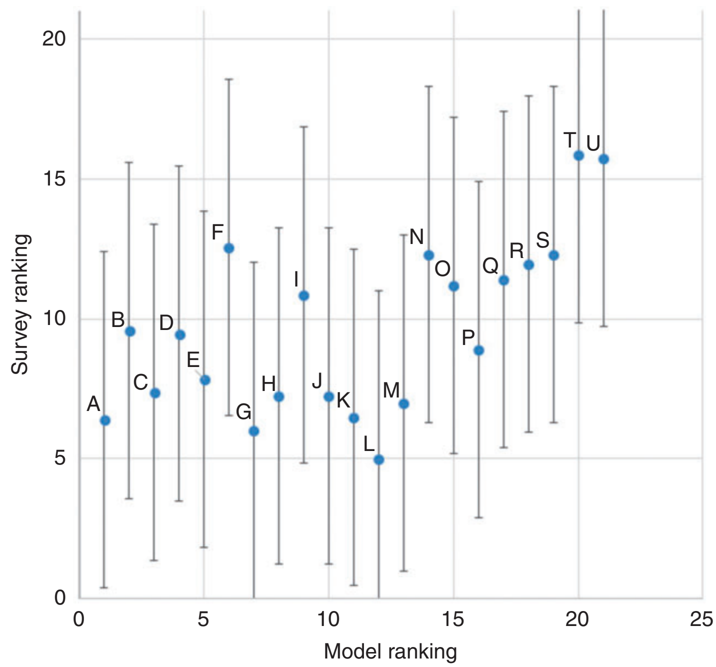

Figure 15. Scatter plot comparing survey results (mean and standard deviation) and simulation model ranking. No correlation is seen.

their experience or greater time since leaving education.

Although the sample size of 108 is not insubstantial, the subpopulations are much smaller (although reasonably sized in social science terms). This suggests a larger experiment with greater statistical power would be worth conducting. However, this analysis does permit some conclusions. There is clearly a great variation in how accurately professional engineers rate model perturbation factors. Of particular note, 25% of the people tested (27 out of 108) performed worse on this task than would be expected if they had

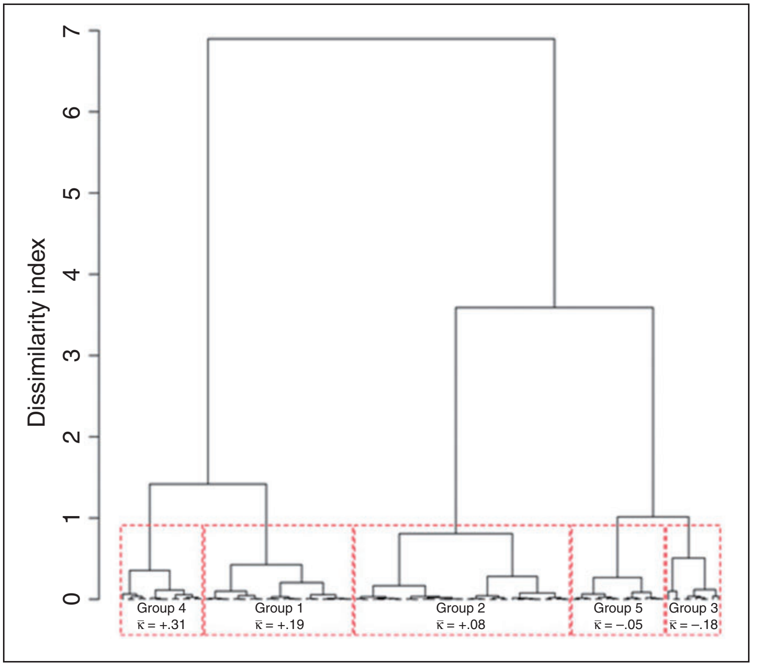

Figure 16. Dendrogram provided by the clustering analysis.

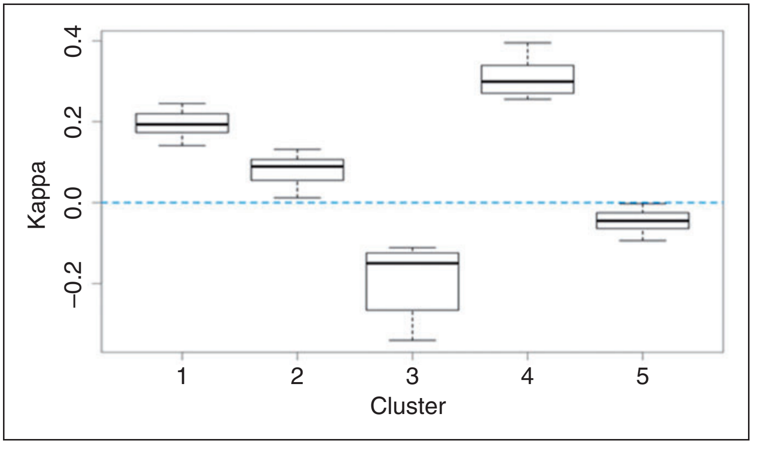

Figure 17. Weighted kappa (i.e. the perturbation judgement skill) distributions for the five groups of participants identified in the cluster analysis.

rated each perturbation factor with a random number between 1 and 5. This suggests that there are some engineers who have systematically skewed ideas about the importance of these perturbation factors. Notably, there are no signs of these people being less experienced or less qualified than their better performing peers.

Table 5. The makeup and performance of the groups identified, as well as the Kappa values of the subpopulations.

Group	Rate	Performance
1	Second rate	Mostly masters level, mostly relatively inexperienced. They do quite well on the task.
2	Third rate	More experienced and more qualified than group 1, this group are nevertheless less skilled on the task.
3	Worst performance	This group do somewhat worse than guessing on the task. Mostly masters level, but qualified some time ago.
4	First rate	This group do well on the task and their ratings show more agreement with the true order of the items than any other group. Notably, even here the ratings are still far from the theoretical maximum of $\kappa = 1.00$. Predominantly masters educated (11 out of 15) with all levels of experience represented.
5	Fourth rate	This group are a lot like group 3, but not quite as egregious. Like group 3, they tend to be experienced and highly qualified.

The kappa values of the subpopulations (mean κ)

Academic level			Years of experience				
Bachelor	Master's	PhD	<1 Years	1–3 Years	3–5 Years	6–10 Years	>10 Years
+0.10	+0.11	−0.01	+0.11	+0.15	+0.06	+0.11	+0.08

Table 6. Performance (in terms of κ) as a function of education and years of experience.

	<1 Year exp.	1–3 Years exp.	3–5 Years exp.	6–10 Years exp.	>10 Years exp.
Bachelors	.05	.13	.12	.19	.05
Masters	.20	.17	.03	.08	.12
PhD	–	–	–	–	−.01

Summary and conclusion

The performance gap is a problem that might affect all new buildings or the refurbishment of older ones. Its existence creates a gap between reality and the policies enacted by governments to reduce energy use and greenhouse gas emissions. Previous studies tried to tackle this problem from various perspectives such as highlighting issues concerned with the role of poor workmanship or occupants' behaviour. The research reported here tackled this problem from the earlier stage of energy modelling, or, more precisely, the building physics literacy of building energy modellers. The literature indicates that this is an understudied area and is highly important as architects, engineers and modellers do not tend to consider themselves as a contributing factor to the performance gap, but rather consider construction quality and occupants to be the problem.

The methodology was chosen specifically to allow a mixed building physics and social science approach, as such the sample size of 108 is particular large. One limitation of the work is the form of the building (a dwelling), and it maybe that a more complex building might have produced even more diversity in the thoughts of the participants and therefore in their scores.

The results are in line with those found by Guyon and Gilles[35] who asked 12 modellers to create a thermal model (using the same software, and in which they were knowledgeable) of a dwelling. A factor of 2.4 was found between the lowest and highest annual heating energy use predictions of the resultant models, and a +18% to

−50% error compared to a validated model of the building. Interestingly, the most experienced, including consultant engineers, performed the worst and had the most diverse performance.

Williamson[36] comments that the results of any simulation will in part be dependent of the philosophy of the modeller and particularly on their ontological views and epistemological beliefs, but that many modellers might not realise this due to their largely positivist position. It would seem reasonable to surmise that this arises out of their positivist-centred educational history. It is quite possible that as a group there is too much belief that a simulation is a true reflection of reality, even if the simulation does not contain a full description of the problem; i.e. modellers might be more concerned about the technical details and accuracy of the simulation engine, than about how their methodology unambiguously captures the problem.

From the results reported here, it is clear that all three tests of literacy suggested in Literacy section have been failed by the sample of participants. Participants do not: (1) approximately agree on the important parameters that need to be included in the model; or (2) approximately agree on the rank order of the importance of a list of possible input parameters; or (3) cannot rank order the impact of given changes to the values of 21 common parameters such that they approximately agree with that given by a sensitivity analysis of the parameters within an industry standard and experimentally validated thermal model of the same building.

Being that the sample size was reasonably large (108), this conclusion is likely to be valid on average also for the whole population of thermal modellers. Future research should therefore identify new ways to teach building physics in both academic and industrial settings, as this work indicates a gap that can be bridged.

The most successful subpopulation shown in Table 6 are those with very recent relevant masters degrees. It is likely that many of these participants, and unlike those graduating before, sat Masters Programmes that contained a large thermal modelling component. It therefore seems reasonable to conclude that this provision should be expanded. However, it is clear that even this subpopulation have $\kappa \ll 1$ and hence those teaching such courses need to face some stark realities and improve their provision. Another possibility is that the culture within engineering consultancy undermines some of the cautionary messages received by engineers during their education, and that because thermal modellers rarely compare their results with the performance of the finished building, there is little feedback or learning, and their personal performance might drift over time. This would give, as observed, a diversity of views about the importance of the various driving parameters.

Acknowledgements

We would like to thank all respondents who participated in the survey conducted in this research and express our appreciation for their valuable comments. All data created during this research are openly available from the University of Bath data archive at http://doi.org/10.15125/BATH-00221

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was supported by the EPSRC funded ENLITEN (EP/K002724/1) and COLBE (EP/M021890/1) projects.

References

1. Perez-Lombard L, Ortiz J and Pout C. A review on buildings energy consumptioninformation. *Energy Build* 2008; 40: 394–398.
2. Menezes AC, et al. Predicted vs. actual performance of non-domestic buildings: using post-occupancy evaluation data to reduce the performance gap. *Appl Energy* 2012; 97: 355–364.
3. Johnston D, Farmer D, Brooke-Peat M, et al. Bridging the domestic building fabric performance gap. *Build Res Inform* 2014; 44: 147–159.

4. de Wilde P. The gap between predicted and measured energy performance of buildings: a framework for investigation. *Autom Constr* 2014; 41: 40–49.
5. Zero Carbon Hub. Closing the gap: design and as-built performance (Evidence review report). London: Zero Carbon Hub, 2014.
6. van Knippenberg D, et al. Organizational identification after a merger: a social identity perspective. *Br J Soc Psychol* 2002; 41: 233–252.
7. Sani F. When subgroups secede: extending and refining the social psychological model of schisms in groups. *Pers Soc Psychol Bull* vol. 31, no. 8, pp. 1074–1086, 2005.
8. Morton TA, Postmes T and Jetten J. Playing the game: when group success is more important than downgrading deviants. *Eur J Soc Psychol* 2007; 37: 599–616.
9. UNESCO. *The Plurality of literacy and its implications for policies and programs*. UNESCO Education Sector Position Paper, vol. 13, 2003.
10. Kress G. *Literacy in the new media age*. New York: Routledge, 2003.
11. Wagner DA. *Monitoring and measuring literacy*. New York: UNESCO, 2006.
12. Attia S, et al. Simulation-based decision support tool for early stages of zero-energy building design. *Energy Build* 2012; 49: 2–15.
13. Hand JW and Crawley DB. *Forget the tool when training new simulation users*. Prague, Czech Republic: IBPSA, 1997.
14. Morbitzer C, et al. *Integration of building simulation into the design process of an architecture practice*. Rio de Janeiro, Brazil: IBPSA, 2001.
15. Augenbroe G. Trends in building simulation. *Build Environ* 2002; 37: 891–902.
16. Mahdavi A, et al. *An inquiry into building performance simulation tools usage by architects in Austria*. Eindhoven: IBPSA, 2003.
17. Ibarra DI and Reinhart CF. *How close do simulation beginners 'really' get?* Glasgow: BS2009, 2009.
18. Schlueter A and Thesseling F. Building information model based energy/exergy performance assessment in early design stages. *Autom Constr* 2009; 18: 153–163.
19. Lam KP, et al. *Energy modelling tools assessment for early design phase*. Pittsburgh, PA: Carnegie Mellon University, 2004.
20. Gernot R and Butler T. Simulation space. In: *Architecture 'in computro' – Integrating Methods and Techniques: 26th eCAADe Conference Proceedings*, eCAADe: Conferences. Antwerpen, Belgium: The Higher Institute of Architectural Sciences, Henry van de Velde, 2008, pp.133–142.
21. De Herde A, et al. *Architect friendly: a comparison of ten different simulation tools*. Glasgow, UK: Building Performance Association, 2009.
22. Weytjens A, et al. *A comparative study of the architect-friendliness of six building performance simulation tools*. Maastricht, the Netherlands: Sustainable Building CIB, 2010.
23. Tianzhen H, Chou SK and Bong TY. Building simulation: an overview of developments and information sources. *Build Environ* 2000; 35: 347–361.
24. Carbon Trust. *Closing the gap: lessons learned on realising the potential of low carbon building design*. London: Carbon Trust, 2011.
25. Newsham GR, Mancini S and Birt B. Do LEED-certified buildings save energy? Yes, but... *Energy Build* 2009; 41: 897–905.
26. Dwyer T. Knowledge is power: benchmarking and prediction of building energy consumption. *Build Serv Eng Res Technol* 2012; 34: 5–7.
27. Almeida N, et al. A framework for combining risk management and performance-based building approaches. *Build Res Inform* 2010; 38: 157–174.
28. Haldi F and Robinson D. On the behaviour and adaptation of office occupants. *Build Environ* 2008; 43: 2163–2177.
29. Korjenic A and Bednar T. Validation and evaluation of total energy use in office buildings: a case study. *Autom Constr* 2012; 23: 64–70.
30. Owen, MS (ed). ASHRAE Handbook – Fundamentals. *ASHRAE data for internal heat gains calculations* (*Chapter 29*). Atlanta, GA: ASHRAE, 2013.
31. Gabe-Thomas E, Walker I and Verplanken B. *Exploring mental representations of home energy practices and habitual energy consumption*. Bath, UK: University of Bath, 2015.
32. Holt N and Walker I. *Research with people "theory, plans and practicals"*. Bath, UK: Palgrave Macmillan, 2009.
33. Bryman A. *Social research methods*. Oxford: Oxford University Press, 2013.
34. Part L. *Conservation of fuel and power*. London: HM Government, 2010.
35. Guyon G. Role of the model user in results obtained from simulation software program. In: *Building simulation conference*, Prague, Czech Republic: IBPSA, September 1997, pp. 8–10.
36. Williamson TJ. Predicting building performance: the ethics of computer simulation. *Build Res Inform* 2010; 38: 401–410.

Appendix 1. The online questionnaire

Survey on how the UK construction industry uses thermal models

Introduction. The following questionnaire is part of research by the Department of Architecture and Civil Engineering at the University of Bath. **We estimate that the survey will take less than 15 minutes to complete.**

The survey aims to make sense of how the construction industry uses thermal models in the UK and how the use of such models might be improved. We know that thermal modellers often have to use their judgement with respect to the time available to model a building and also have to produce models before all the architectural details are known. We would like to know how you make these judgements with respect to which parameters to include accurately or which you might not be overly concerned about if only an approximate value was available. For example, we would like to know, **given the building detailed below**, do you consider it is more important to know details of the positions of the windows in the walls, or when people occupy the building?

We do not ask for any personally identifiable information (such as names, date of birth etc.). The data we collect from you will be converted into a generic form profile and will be used only for research purposes. Please don't think overly carefully about your answers, we want to know how you normally work in practice and what your natural thoughts are. This is not a test!

General information

Q1: Please indicate your years of experience in the construction industry.

- Less than 1 year (graduate)/1–3 years/3–5 years/6–10 years/over 10 years.

Q2: Please indicate the highest degree you have received (related to the construction industry).

- Bachelor degree/Master's degree/PhD degree/other (please specify).

Q3: Please indicate the simulation software(s) that you use for energy analysis.

- IES VE/TAS/Design Builder/Energy Plus/PHPP/eQuest/other (please specify).

Case study description

Note: Questions shown in next pages are related to the case study shown below.

Below you can see the ground and first floor plans (Figure 18) as well as the construction details (Figure 19) of a house located in Exeter, UK. Both exterior and location map views were captured from Google maps and shown in Figures 20 and 21. Although a dynamic simulation would not normally be used on such a building, we have chosen this as it is a relatively simple case.

General information

House type: Semi-detached.
Stories: 2 (No basement).
Internal floor area: 80 m^2.
Glazing type: Double glazed.
Location: Exeter, UK.

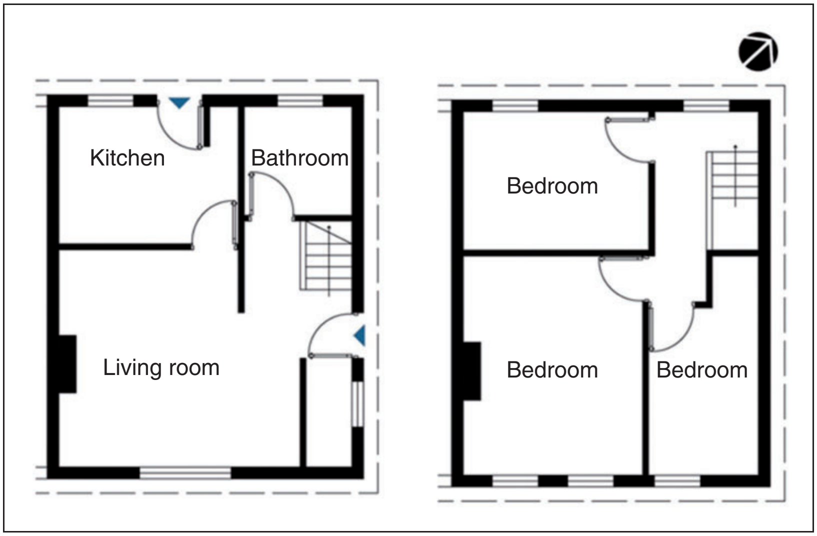

Figure 18. Ground and first floor plans for the case study dwelling.

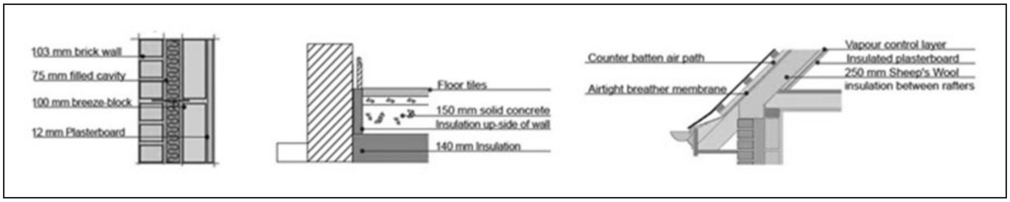

Figure 19. Construction details for walls (left), ground floor (middle) and roof (right).

Figure 20. Exterior view of the case study building from the South-East facade (Taken from the EPSRC funded ENLITEN (EP/K002724/1) project team at the University of Bath, UK). Image taken from Google maps.

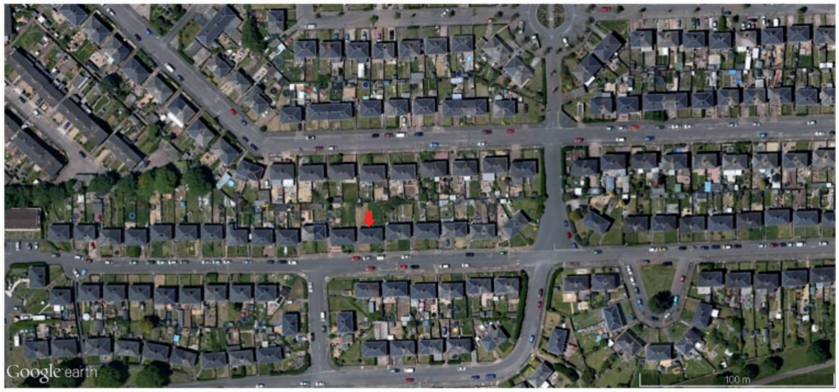

Figure 21. The location of the building. The red arrow is pointing at the chosen case study building. Image taken from Google maps.

Free-form questions

Q4: Please list below the three most important parameters that if not included or included less accurately in a thermal model of this building (shown above) might affect the annual heating demand significantly.

Q5: Please list below the three parameters that you might not normally include in a thermal model of this building (shown above) as they do not have a great impact on the annual heating demand.

Q6: Please list below any other parameters that you might include in a thermal model of this building (shown above) and might have a moderate effect on the annual heating demand.

Given list

In the following final question, we are aiming to identify the relationship between the annual heating energy use predicted by a thermal model of the building and the thoughts of the design team on errors that some parameters might have.

Q 6: For the case study shown, Please rate the list of parameters described below based on your judgement of the impact of differences applied to each parameter (shown below in brackets) on the annual heating demand. These errors might be due to lack of knowledge in the design stage or poor workmanship on site. For example, does a 10% error in the airtightness value have more or less impact than a 20% error in roof U-value? Please indicate the relative size of impact for each parameter by marking them with a scale from 1 to 5.

- Airtightness (20% greater than modelled)
- Internal heat gains from appliances and lighting (10% greater than modelled)
- Windows recessed 100 mm further into the building
- Density of block used as inner leaf of wall (10% greater than modelled)
- Glazing ratio (10% greater than actual ratio)
- Roof U-value (20% greater than modelled value)
- Walls U-value (20% greater than modelled value)

- Ground floor U-value (20% greater than modelled value)
- Installed window U-value (20% greater than modelled value)
- Shading from the surrounding environment (Ignoring the surrounding homes)
- Using internal dimensions for the building rather than external
- Occupancy period (25% greater than modelled period)
- Ventilation (Assuming the air flow is constant at 1 ach when occupied, against the base case of assuming windows are open during occupancy period, if $T_{in} > 25°C$, or RH >75%, or CO_2 concentration >1000 ppm)
- Thermal bridge (Ignoring thermal bridges)
- Winter indoor temperature set-point (The modelled value being 2°C lower than reality)
- Ventilation rate (Assuming 1.1 ach rather than 1 ach)
- The position of windows in the walls (Assuming a 0.5 m vertical shift down from the actual position in each façade)
- Assuming thermostats in each room rather than just in the living room
- Ignoring the use of curtains at night
- Ignoring heat gains from cooking
- Ignoring the fact that the external doors might be opened 10 times a day for 30 seconds each time

Last step

If you wish to know our findings later in the year, please fill in your email address below. Kindly know that your email address will be kept separately from your answers to keep all results anonymous (optional).

Appendix 2. Raw survey results

The following Tables 7 and 8 are indicating the numerical data concerning the results of both the free form and given list questionnaires.

Table 7. Free-form survey responses.

Input parameters	Number of times mentioned	Number of times not mentioned
U-values	108	0
Internal heat gains	104	4
Air tightness	83	25
Ventilation rate	81	27
Shading from surrounding environment	74	34
Glazing type	65	43
Occupants behaviour	62	46
Thermal bridging	58	50
Building orientation	36	72
Thermal mass	34	74
The use of curtains	33	74
Glazing ratio	32	76
Heat loss from system pipes	26	82
Heating system efficiency	25	83
Indoor surfaces colour	20	88
Windows g-value	16	92
Heating system set-point	10	98
Weather data	9	99
Solar heat gains	8	100
Internal doors opening	4	104

Table 8. Given list survey responses.

Input parameter	Weight scale 1	2	3	4	5	Weighted average
Glazing ratio	6	6	9	26	61	4.20
Installed window U-value	9	3	18	27	51	4.00
Walls U-value	6	6	24	27	45	3.92
Occupancy period	9	6	15	34	44	3.91
Airtightness (infiltration rate)	6	12	18	33	39	3.81
Roof U-value	6	3	39	24	36	3.75
Thermal bridging	6	9	27	30	36	3.75
Winter indoor temp. set-point	9	12	12	42	33	3.72
Natural ventilation	15	11	12	31	39	3.63
Ground floor U-value	9	21	24	24	30	3.42
Building geometry	15	15	24	30	24	3.31
Ventilation rate	13	12	35	27	21	3.29
Shading from surroundings	15	33	21	12	27	3.03
Windows recession	12	27	36	21	12	2.94
The position of windows in walls	21	21	33	12	21	2.92
Density of block used as inner leaf of wall	21	27	24	24	12	2.81
IHG from appliances and lighting	15	33	33	15	12	2.78
External doors opening	18	27	36	18	9	2.75
IHG from cooking	39	30	33	6	0	2.06
Thermostats location	27	24	27	15	15	2.69
The use of curtains	45	33	15	12	3	2.03